(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77124 A2

(51) International Patent Classification[7]: C07H

(21) International Application Number: PCT/US01/10670

(22) International Filing Date: 3 April 2001 (03.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/194,362 5 April 2000 (05.04.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BENTIVEGNA, Steven C.** [US/US]; 49 Greenview, Terrace, Middletown, CT 06457 (US). **CHOI, Julie, Y.** [US/US]; 38 Elizabeth Street, West Haven, CT 06516 (US). **KLIEM, Stefanie, E.** [DE/US]; 1298 Hartford Turnpike, Apt. 4E, North Haven, CT 06473 (US). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apt. 11B, North Haven, CT 06473 (US).

(74) Agents: **SHTIVELBAND, Inna** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/77124 A2

(54) Title: HAPLOTYPES OF THE APOA4 GENE

(57) Abstract: Novel single nucleotide polymorphisms in the human apolipoprotein A-IV (APOA4) gene are described. In addition, various genotypes, haplotypes and haplotype pairs for the APOA4 gene that exist in the population are described. Compositions and methods for haplotyping and/or genotyping the APOA4 gene in an individual are also disclosed. Polynucleotides containing one or more of the APOA4 polymorphisms disclosed herein are also described.

HAPLOTYPES OF THE APOA4 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/194,362 filed April 5, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human apolipoprotein A-IV (APOA4) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial

for other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of atherosclerosis and coronary heart disease is the apolipoprotein A-IV (APOA4) gene or its encoded product. APOA4 is a glycoprotein synthesized by the human intestine that acts as a satiety signal (Tso et al., J. Nutr. 1999; 129:1503-1506). APOA4 synthesis is markedly stimulated by fat absorption,the formation of chylomicrons, and by a factor from the ileum, probably peptide tyrosine-tyrosine (PYY). Chronic ingestion of a high fat diet blunts the intestinal APOA4 response to fat intake and may explain why the chronic ingestion of a

high fat diet predisposes both animals and humans to obesity.

APOA4 is also involved in the metabolism of triglyceride- rich lipoproteins and high-density lipoproteins and may thus play an important role in lipid metabolism (Tenkanen and Ehnholm, Ann. Med. 1992; 24:369-374). Duverger et al., (Science 1996; 273:966-968) showed that human APOA4 expressed in the liver of C57BL/6 mice and mice deficient in APOE, both of which are prone to atherosclerosis, caused a substantial reduction in the size of atherosclerotic lesions. Recently, Kronenberg et al., (J. Am. Coll. Cardiol. 2000; 36:751-757) showed that plasma APOA4 levels were significantly lower in 114 male Caucasian subjects with angiographically defined coronary artery disease when compared with 114 age-adjusted male controls, suggesting that low APOA4 concentrations are associated with coronary artery disease and that APOA4 may play an anti-atherogenic role in humans.

The apolipoprotein A-IV gene is located on chromosome 11q23-qter and contains 3 exons that encode a 396 amino acid protein. Reference sequences for the APOA4 gene (Genaissance Reference No. 1765199; SEQ ID NO: 1), coding sequence (GenBank Accession No:NM_000482.2), and protein are shown in Figures 1, 2 and 3, respectively.

APOA4 is genetically polymorphic in humans, in whom two major isoproteins (APOA4 1 and APOA4 2) are present and have differences that influence the APOA4 phenotype in lipid metabolism (Verges, *Diabete Metab* 1995; 21:99-105). A polymorphism of thymine or guanine at a position corresponding to nucleotide 3279 in Figure 1 results in an amino acid variation of histidine or glutamine at a position corresponding to amino acid 380 in Figure 3, with the histidine residue characterizing the APOA4 2 isoprotein (Lohse et al., *J. Biol. Chem.* 1990; 265:10061-10064). A polymorphism of adenine or guanine at a position corresponding to nucleotide 2579 in Figure 1, which results in a variation of asparagine or serine at amino acid 147 in Figure 3 has been identified (Tenkanen et al., *Biochim. Biophys. Acta* 1992; 1138:27-33). A polymorphism of guanine or adenine at a position corresponding to nucleotide 1042 in Figure 1 causes an amino acid variation of valine or methionine at a position corresponding to amino acid 13 in Figure 3 (von Eckardstein et al., *Am. J. Hum. Genet.* 1992; 50:1115-1128). In a family study, the methionine allele was associated with lower HDL-cholesterol and higher low-density lipoprotein-cholesterol concentrations, which are characteristics of atherosclerosis. Deeb et al., (*Human Mutation* 1996;8(4):319-25) described a polymorphism of guanine or thymine at a position corresponding to nucleotide 2620 in Figure 1 that results in an amino acid variation of alanine or serine at a position corresponding to amino acid 161 in Figure 3. A polymorphism of adenine or thymine at a position corresponding to nucleotide 3238 in Figure 1 results in an amino acid variation of threonine or serine at a position corresponding to amino acid 367 in Figure 3 (Wang et al., *Science* 1998; 280:1077-1082). In addition, Halushka et al., (*Nat. Genet.* 1999; 22:239-247) and Cennarro et al., (*Molecular and Cellular Probes* 1998, 12:117-120) identified several other polymorphisms in Figure 1 that do not affect protein translation: guanine or cytosine at a position corresponding to nucleotide 858, guanine or adenine at a position corresponding

to nucleotide 908, adenine or guanine at a position corresponding to nucleotide 1449, thymine or cytosine at a position corresponding to nucleotide 1559, cytosine or thymine at a position corresponding to nucleotide 1700, cytosine or thymine at a position corresponding to nucleotide 2355, thymine or cytosine at a position corresponding to nucleotide 2361, cytosine or thymine at a position corresponding to nucleotide 2589, cytosine or thymine at a position corresponding to nucleotide 2919, cytosine or thymine at a position corresponding to nucleotide 2985, guanine or adenine at a position corresponding to nucleotide 3027, and finally guanine or thymine at a position corresponding to nucleotide 3147.

Because of the potential for variation in the APOA4 gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in the APOA4 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of APOA4 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 10 novel polymorphic sites in the APOA4 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 698 (PS2), 1042 (PS5), 1560 (PS8), 2464 (PS12), 2471 (PS13), 2520 (PS14), 2719 (PS18), 2721 (PS19), 3036 (PS23), and 3196 (PS25). The polymorphisms at these sites are cytosine or thymine at PS2, guanine or adenine at PS5, guanine or adenine at PS8, cytosine or thymine at PS12, cytosine or thymine at PS13, cytosine or adenine at PS14, guanine or adenine at PS18, guanine or adenine at PS19, guanine or adenine at PS23, and thymine or guanine at PS25. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified sites at nucleotide positions 571 (PS1), 858 (PS3), 908 (PS4), 1449 (PS6), 1559 (PS7), 1700 (PS9), 2355 (PS10), 2361 (PS11), 2579 (PS15), 2589 (PS16), 2620 (PS17), 2919 (PS20), 2985 (PS21), 3027 (PS22), 3147 (PS24), 3238 (PS26) and 3279 (PS27), in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-27 in the APOA4 gene, which are shown below in Tables 4 and 3, respectively. Each of these APOA4 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the APOA4 gene that exists in the human population.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the APOA4 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25 in both copies of the APOA4 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe

or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel APOA4 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel APOA4 polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-27. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the APOA4 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the APOA4 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's APOA4 gene is defined by one of the APOA4 haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's APOA4 gene are defined by one of the APOA4 haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. The method for establishing the APOA4 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with APOA4 activity, e.g., atherosclerosis and coronary heart disease.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate APOA4 as a candidate target for treating a specific condition or disease predicted to be associated with APOA4 activity. Determining for a particular population the frequency of one or more of the individual APOA4 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue APOA4 as a target for treating the specific disease of interest. In particular, if variable APOA4 activity is associated with the disease, then one or more APOA4 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed APOA4 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable APOA4 activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any APOA4 haplotype or haplotype pair, apply the information derived from detecting APOA4 haplotypes in an individual to decide whether modulating APOA4 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting APOA4 to treat a specific condition or disease predicted to be associated with APOA4 activity. For example, detecting which of the APOA4 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent APOA4 isoforms present in the disease population. Thus, without requiring any *a priori*

knowledge of the phenotypic effect of any particular APOA4 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the APOA4 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with APOA4 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the APOA4 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute APOA4 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of an APOA4 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any APOA4 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and an APOA4 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the APOA4 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the APOA4 genotype or haplotype in a reference population. A higher frequency of the APOA4 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the APOA4 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the APOA4 haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for atherosclerosis and coronary heart disease.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the APOA4 gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at PS2, adenine at PS5, adenine at PS8, thymine at PS12, thymine at PS13, adenine at PS14, adenine at PS18, adenine at PS19, adenine at PS23, and guanine at PS25. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of adenine at PS1, cytosine at PS3, adenine at PS4, guanine at PS6, cytosine at PS7, thymine at PS9, thymine at PS10, cytosine at PS11, guanine at PS15, thymine at PS16, thymine at PS17, thymine at PS20, thymine at PS21, adenine at PS22, thymine at PS24, thymine at PS26 and guanine at PS27.

A particularly preferred polymorphic variant is an isogene of the APOA4 gene. An APOA4 isogene of the invention comprises guanine or adenine at PS1, cytosine or thymine at PS2, guanine or

cytosine at PS3, guanine or adenine at PS4, guanine or adenine at PS5, adenine or guanine at PS6, thymine or cytosine at PS7, guanine or adenine at PS8, cytosine or thymine at PS9, cytosine or thymine at PS10, thymine or cytosine at PS11, cytosine or thymine at PS12, cytosine or thymine at PS13, cytosine or adenine at PS14, adenine or guanine at PS15, cytosine or thymine at PS16, guanine or thymine at PS17, guanine or adenine at PS18, guanine or adenine at PS19, cytosine or thymine at PS20, cytosine or thymine at PS21, guanine or adenine at PS22, guanine or adenine at PS23, guanine or thymine at PS24, thymine or guanine at PS25, adenine or thymine at PS26 and thymine or guanine at PS27. The invention also provides a collection of APOA4 isogenes, referred to herein as an APOA4 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for an APOA4 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of adenine at a position corresponding to nucleotide 37, thymine at a position corresponding to nucleotide 325, thymine at a position corresponding to nucleotide 332, adenine at a position corresponding to nucleotide 381, adenine at a position corresponding to nucleotide 580, adenine at a position corresponding to nucleotide 582, adenine at a position corresponding to nucleotide 897, and guanine at a position corresponding to nucleotide 1057. In a preferred embodiment, the polymorphic variant comprises one or more additional polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 87, thymine at a position corresponding to nucleotide 216, cytosine at a position corresponding to nucleotide 222, guanine at a position corresponding to nucleotide 440, thymine at a position corresponding to nucleotide 450, thymine at a position corresponding to nucleotide 481, thymine at a position corresponding to nucleotide 780, thymine at a position corresponding to nucleotide 846, adenine at a position corresponding to nucleotide 888, thymine at a position corresponding to nucleotide 1008, thymine at a position corresponding to nucleotide 1099 and guanine at a position corresponding to nucleotide 1140. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an APOA4 isogene defined by haplotypes 1-36. .

Polynucleotides complementary to these APOA4 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the APOA4-gene will be useful in studying the expression and function of APOA4, and in expressing APOA4 protein for use in screening for candidate drugs to treat diseases related to APOA4 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express APOA4 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the APOA4 protein. The reference amino acid

sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of methionine at a position corresponding to amino acid position 13, valine at a position corresponding to amino acid position 111, lysine at a position corresponding to amino acid position 127, methionine or isoleucine at a position corresponding to amino acid position 194, and alanine at a position corresponding to amino acid position 353. In some embodiments, the polymorphic variant also comprises at least one variant amino acid selected from the group consisting of serine at a position corresponding to amino acid position 147, serine at a position corresponding to amino acid position 161, serine at a position corresponding to amino acid position 367 and glutamine at a position corresponding to amino acid position 380. A polymorphic variant of APOA4 is useful in studying the effect of the variation on the biological activity of APOA4 as well as on the binding affinity of candidate drugs targeting APOA4 for the treatment of atherosclerosis and coronary heart disease.

The present invention also provides antibodies that recognize and bind to the above polymorphic APOA4 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the APOA4 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the APOA4 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against APOA4 protein, and for testing the efficacy of therapeutic agents and compounds for atherosclerosis and coronary heart disease in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the APOA4 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the APOA4 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing APOA4 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the APOA4 gene (Genaissance Reference No. 1765199; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:56 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25).

Figure 2 illustrates a reference sequence for the APOA4 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the APOA4 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the APOA4 gene. As described in more detail below, the inventors herein discovered 36 isogenes of the APOA4 gene by characterizing the APOA4 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (CA) (22 individuals), African descent (AF) (20 individuals), Asian (AS) (20 individuals), or Hispanic/Latino (HL) (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (AM) (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The APOA4 isogenes present in the human reference population are defined by haplotypes for 27 polymorphic sites in the APOA4 gene, 10 of which are believed to be novel. The APOA4 polymorphic sites identified by the inventors are referred to as PS1-27 to designate the order in which they are located in the gene (see Table 2 below), with the novel polymorphic sites referred to as PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25. Using the genotypes identified in the Index Repository for PS1-27 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the APOA4 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the APOA4 gene include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether APOA4 is a suitable target for drugs to treat atherosclerosis and coronary heart disease, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular

nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with

the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the APOA4 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel APOA4 polymorphisms and haplotypes identified herein.

The compositions comprise at least one APOA4 genotyping oligonucleotide. In one embodiment, an APOA4 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of an APOA4 polynucleotide, i.e., an APOA4 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-APOA4 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the APOA4 gene using the polymorphism information provided herein in conjunction with the known sequence information for the APOA4 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of

the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting APOA4 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GCCACCAYTTCTAAC (SEQ ID NO:4) and its complement,
GGCCCTGRTGGCTGT (SEQ ID NO:5) and its complement,
AGTGTGTRGTGGTGA (SEQ ID NO:6) and its complement,
GGAGGAGYTGAGGGC (SEQ ID NO:7) and its complement,
CTGAGGGYCCGGCTG (SEQ ID NO:8) and its complement,
GGGACAAMCTGCGAG (SEQ ID NO:9) and its complement,
CCAGAACRTGGAGGA (SEQ ID NO:10) and its complement,
AGAACGTRGAGGAGC (SEQ ID NO:11) and its complement,
ACCAGCARGTGGAGG (SEQ ID NO:12) and its complement, and
GGTCAACKCCTTCTT (SEQ ID NO:13) and its complement.

A preferred ASO primer for detecting APOA4 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GGAACAGCCACCAYT (SEQ ID NO:14); CGATAGGTTAGAART (SEQ ID NO:15);
GACCCTGGCCCTGRT (SEQ ID NO:16); CCGGCGACAGCCAYC (SEQ ID NO:17);
GACTACAGTGTGTRG (SEQ ID NO:18); TCCCCGTCACCACYA (SEQ ID NO:19);
GGAGCTGGAGGAGYT (SEQ ID NO:20); AGCCGGGCCCTCARC (SEQ ID NO:21);
GAGGAGCTGAGGGYC (SEQ ID NO:22); GGGCAGCAGCCGGRC (SEQ ID NO:23);
AGATCGGGGACAAMC (SEQ ID NO:24); GAAGCTCTCGCAGKT (SEQ ID NO:25);
GATCGACCAGAACRT (SEQ ID NO:26); TTGAGCTCCTCCAYG (SEQ ID NO:27);
TCGACCAGAACGTRG (SEQ ID NO:28); CCTTGAGCTCCTCYA (SEQ ID NO:29);
ACCTGGACCAGCARG (SEQ ID NO:30); GGAACTCCTCCACYT (SEQ ID NO:31);
GGACAAGGTCAACKC (SEQ ID NO:32); and GTGCTGAAGAAGGMG (SEQ ID NO:33).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting APOA4 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

ACAGCCACCA (SEQ ID NO:34); TAGGTTAGAA (SEQ ID NO:35);
CCTGGCCCTG (SEQ ID NO:36); GCGACAGCCA (SEQ ID NO:37);
TACAGTGTGT (SEQ ID NO:38); CCGTCACCAC (SEQ ID NO:39);
GCTGGAGGAG (SEQ ID NO:40); CGGGCCCTCA (SEQ ID NO:41);
GAGCTGAGGG (SEQ ID NO:42); CAGCAGCCGG (SEQ ID NO:43);
TCGGGGACAA (SEQ ID NO:44); GCTCTCGCAG (SEQ ID NO:45);
CGACCAGAAC (SEQ ID NO:46); AGCTCCTCCA (SEQ ID NO:47);
ACCAGAACGT (SEQ ID NO:48); TGAGCTCCTC (SEQ ID NO:49);
TGGACCAGCA (SEQ ID NO:50); ACTCCTCCAC (SEQ ID NO:51);
CAAGGTCAAC (SEQ ID NO:52); and CTGAAGAAGG (SEQ ID NO:53).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

APOA4 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized APOA4 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the APOA4 gene in an individual. As used herein, the terms "APOA4 genotype" and "APOA4 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the APOA4 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the APOA4 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25 in the two copies to assign an APOA4 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a preferred embodiment of the genotyping method, the identity of the nucleotide pair at one or more of the polymorphic sites selected from the group consisting of PS1, PS3, PS4, PS6, PS7, PS9, PS10, PS11, PS15, PS16, PS17, PS20, PS21, PS22, PS24, PS26, and PS27 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-27.

Typically, the nucleic acid sample is isolated from a biological sample taken from the

individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the APOA4 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions. If an APOA4 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the APOA4 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25 in that copy to assign an APOA4 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the APOA4 gene or fragment such as one of the methods described above for preparing APOA4 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two APOA4 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional APOA4 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the APOA4 gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4, PS6, PS7, PS9, PS10, PS11, PS15, PS16, PS17, PS20, PS21, PS22, PS24, PS26, and PS27. In a particularly preferred embodiment, the nucleotide at each of PS1-27 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the APOA4 haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual's APOA4 gene, the phased sequence of nucleotides present at each of PS1-27. The present invention also contemplates that typically only a subset of PS1-27 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, an APOA4 haplotype pair is determined for an individual by

identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25 in each copy of the APOA4 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-27 in each copy of the APOA4 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the APOA4 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be

used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the APOA4 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798.

Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's APOA4 haplotype pair is predicted from its APOA4 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying an APOA4 genotype for the individual at two or more APOA4 polymorphic sites described herein, emmerating all possible haplotype pairs which are consistent with the genotype, accessing data containing APOA4 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the APOA4 haplotype pairs shown in Table 3.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland,

MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1/H_2 is equal to $p_{H-W}(H_1/H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1/H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting an APOA4 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*).

The invention also provides a method for determining the frequency of an APOA4 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel APOA4 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for APOA4 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an

association between a trait and an APOA4 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular APOA4 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that APOA4 genotype, haplotype or haplotype pair. Preferably, the APOA4 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting APOA4 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and an APOA4 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the

existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the APOA4 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and APOA4 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their APOA4 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the APOA4 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between APOA4 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or

local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the APOA4 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of APOA4 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the APOA4 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the APOA4 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying APOA4 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the APOA4 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant APOA4 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25, and may also comprise one or more additional polymorphisms selected from the group consisting of adenine at PS1, cytosine at PS3, adenine at PS4, guanine at PS6, cytosine at PS7, thymine at PS9, thymine at PS10, cytosine at PS11, guanine at PS15, thymine at PS16, thymine at PS17, thymine at PS20, thymine at PS21, adenine at PS22, thymine at PS24, thymine at PS26 and guanine at PS27. Similarly, the nucleotide sequence of a variant fragment of the APOA4 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the APOA4 gene (or other reported APOA4 sequences) or to portions of the reference sequence (or other reported APOA4

sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of thymine at PS2, adenine at PS5, adenine at PS8, thymine at PS12, thymine at PS13, adenine at PS14, adenine at PS18, adenine at PS19, adenine at PS23, and guanine at PS25. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the APOA4 gene which is defined by any one of haplotypes 1-36 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the APOA4 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

APOA4 isogenes may be isolated using any method that allows separation of the two "copies" of the APOA4 gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides APOA4 genome anthologies, which are collections of APOA4 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. An APOA4 genome anthology may comprise individual APOA4 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the APOA4 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred APOA4 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded APOA4 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other

regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant APOA4 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the APOA4 gene will produce APOA4 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of an APOA4 cDNA comprising a nucleotide sequence which is a polymorphic variant of the APOA4 reference coding sequence shown in Figure 2. Thus, the invention also provides APOA4 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of adenine at a position corresponding to nucleotide 37, thymine at a position corresponding to nucleotide 325, thymine at a position corresponding to nucleotide 332, adenine at a position corresponding to nucleotide 381, adenine at a position corresponding to nucleotide 580, adenine at a position corresponding to nucleotide 582, adenine at a position corresponding to nucleotide 897, and guanine at a position corresponding to nucleotide 1057,and may also comprise one or more additional polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 87, thymine at a position corresponding to nucleotide 216, cytosine at a position corresponding to nucleotide 222, guanine at a position corresponding to nucleotide 440, thymine at a position corresponding to nucleotide 450, thymine at a position corresponding to

nucleotide 481, thymine at a position corresponding to nucleotide 780, thymine at a position corresponding to nucleotide 846, adenine at a position corresponding to nucleotide 888, thymine at a position corresponding to nucleotide 1008, thymine at a position corresponding to nucleotide 1099 and guanine at a position corresponding to nucleotide 1140. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an APOA4 isogene defined by haplotypes 1-36. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized APOA4 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of an APOA4 gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the APOA4 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the APOA4 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the APOA4 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular APOA4 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the APOA4 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular APOA4 isogene. Expression of an APOA4 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions −10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I.

Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of APOA4 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of APOA4 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference APOA4 amino acid sequence shown in Figure 3. The location of a variant amino acid in an APOA4 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). An APOA4 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of methionine at a position corresponding to amino acid position 13, valine at a position corresponding to amino acid position 111, lysine at a position corresponding to amino acid position 127, methionine or isoleucine at a position corresponding to amino acid position 194, and alanine at a position corresponding to amino acid position 353, and may also comprise one or more additional variant amino acids selected from the group consisting of serine at a position corresponding to amino acid position 147, serine at a position corresponding to amino acid position 161, serine at a position corresponding to amino acid position 367 and glutamine at a position corresponding to amino acid position 380. The invention specifically excludes amino acid sequences identical to those previously identified for APOA4, including SEQ ID NO:3, and previously described fragments thereof. APOA4 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having any combination of amino acid variations described herein.

The invention also includes APOA4 peptide variants, which are any fragments of an APOA4 protein variant that contain one or more of the amino acid variations describwed herein. An APOA4 peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such APOA4 peptide variants may be useful as antigens to generate antibodies specific for one of the above APOA4 isoforms. In addition, the APOA4 peptide variants may be useful in drug screening assays.

An APOA4 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant APOA4 genomic and cDNA sequences as described above.

Alternatively, the APOA4 protein variant may be isolated from a biological sample of an individual having an APOA4 isogene which encodes the variant protein. Where the sample contains two different APOA4 isoforms (i.e., the individual has different APOA4 isogenes), a particular APOA4 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular APOA4 isoform but does not bind to the other APOA4 isoform.

The expressed or isolated APOA4 protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the APOA4 protein as discussed further below. APOA4 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant APOA4 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric APOA4 protein. The non-APOA4 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the APOA4 and non-APOA4 portions so that the APOA4 protein may be cleaved and purified away from the non-APOA4 portion.

An additional embodiment of the invention relates to using a novel APOA4 protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known APOA4 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The APOA4 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to an APOA4 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the APOA4 protein(s) of interest and then washed. Bound APOA4 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel APOA4 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the APOA4 protein.

In yet another embodiment, when a particular APOA4 haplotype or group of APOA4 haplotypes encodes an APOA4 protein variant with an amino acid sequence distinct from that of APOA4 protein isoforms encoded by other APOA4 haplotypes, then detection of that particular APOA4 haplotype or group of APOA4 haplotypes may be accomplished by detecting expression of the encoded APOA4 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel APOA4 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The APOA4 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the APOA4 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the APOA4 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the APOA4 protein variant from solution as well as react with APOA4 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect APOA4 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel APOA4 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the APOA4 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in

Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of APOA4 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the APOA4 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into APOA4 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired APOA4 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the APOA4 isogene is introduced into a cell in such a way that it recombines with the endogenous APOA4 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired APOA4 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the APOA4 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the APOA4 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant APOA4 gene

are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the APOA4 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human APOA4 isogene and producing human APOA4 protein can be used as biological models for studying diseases related to abnormal APOA4 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel APOA4 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel APOA4 isogenes; an antisense oligonucleotide directed against one of the novel APOA4 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel APOA4 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel APOA4 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient

and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the APOA4 gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The APOA4 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the APOA4 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the APOA4 gene were amplified using either PCR primer pairs represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1, or 'tailed' PCR primers, each of which includes a universal sequence forming a noncomplementary 'tail' attached to the 5′ end of each unique sequence in the PCR primer pairs. The universal 'tail' sequence for the forward PCR primers comprises the sequence 5′-TGTAAAACGACGGCCAGT-3′ (SEQ ID NO:54) and the universal 'tail' sequence for the reverse PCR primers comprises the sequence 5′-AGGAAACAGCTATGACCAT-3′ (SEQ ID NO:55). The nucleotide positions of the first and last nucleotide of the forward and reverse primers for each region amplified are presented below and correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	495-517	complement of 1086-1065	592 nt
Fragment 2	749-771	complement of 1310-1289	562 nt
Fragment 3	798-819	complement of 1347-1326	550 nt
Fragment 4	1128-1150	complement of 1723-1701	596 nt
Fragment 5	2142-2164	complement of 2733-2712	592 nt
Fragment 6	2415-2437	complement of 3015-2994	601 nt
Fragment 7	2699-2719	complement of 3212-3189	514 nt
Fragment 8	2966-2985	complement of 3520-3500	555 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:

97°C - 2 min. 1 cycle

$$\left.\begin{array}{l}97^{\circ}\mathrm{C} - 15\ \mathrm{sec.}\\ 70^{\circ}\mathrm{C} - 45\ \mathrm{sec.}\\ 72^{\circ}\mathrm{C} - 45\ \mathrm{sec.}\end{array}\right\}\quad 10\ \mathrm{cycles}$$

$$\left.\begin{array}{l}97^{\circ}\mathrm{C} - 15\ \mathrm{sec.}\\ 64^{\circ}\mathrm{C} - 45\ \mathrm{sec.}\\ 72^{\circ}\mathrm{C} - 45\ \mathrm{sec.}\end{array}\right\}\quad 35\ \mathrm{cycles}$$

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using either the primer sets represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1, or the appropriate universal 'tail' sequence as a primer. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	Tailed Sequence	Tailed Sequence
Fragment 2	797-816	complement of 1185-1166
Fragment 3	869-888	complement of 1344-1324
Fragment 4	1292-1312	complement of 1659-1640
Fragment 5	2193-2214	complement of 2653-2635
Fragment 6	Tailed Sequence	Tailed Sequence
Fragment 7	2724-2743	complement of 3187-3167
Fragment 8	2991-3011	complement of 3456-3437

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res*. 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the APOA4 gene are listed in Table 2 below.

Table 2. Polymorphic Sites Identified in the APOA4 Gene

Polymorphic Site Number	PolyId	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1[R]	7652349	571	G	A		
PS2	7652255	698	C	T		
PS3[R]	1765218	858	G	C		
PS4[R]	1765223	908	G	A		
PS5	1765230	1042	G	A	37	V13M
PS6[R]	1765240	1449	A	G	87	T29T
PS7[R]	1765243	1559	T	C		
PS8	1765245	1560	G	A		
PS9[R]	1765247	1700	C	T		
PS10[R]	1765251	2355	C	T	216	Y72Y
PS11[R]	1765253	2361	T	C	222	G74G
PS12	1765256	2464	C	T	325	L109L
PS13	7649848	2471	C	T	332	A111V
PS14	7649945	2520	C	A	381	N127K
PS15[R]	1765260	2579	A	G	440	N147S
PS16[R]	7650041	2589	C	T	450	A150A
PS17[R]	7650137	2620	G	T	481	A161S
PS18	7650233	2719	G	A	580	V194I
PS19	1765264	2721	G	A	582	V194I
PS20[R]	1765266	2919	C	T	780	A260A
PS21[R]	1765270	2985	C	T	846	T282T
PS22[R]	1765274	3027	G	A	888	L296L
PS23	1765278	3036	G	A	897	Q299Q
PS24[R]	1765282	3147	G	T	1008	V336V
PS25	1765285	3196	T	G	1057	S353A
PS26[R]	1765287	3238	A	T	1099	T367S
PS27[R]	1765295	3279	T	G	1140	H380Q

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.
[R]Reported in the literature

EXAMPLE 2

This example illustrates analysis of the APOA4 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 3. Genotypes and Haplotype Pairs Observed in the APOA4 Gene

Genotype	Polymorphic Sites																											HAP	
Number	PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16	PS 17	PS 18	PS 19	PS 20	PS 21	PS 22	PS 23	PS 24	PS 25	PS 26	PS 27	PAIR	
1	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	1
2	G	C	G	G	G	A/G	C	G	C	C	T	C	C	C	A/G	C	G	G	G	C	C	G	G	G	T	A	G	1	2
3	G	C	G	G	G	A	C/T	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	3
4	G	C	G	G	G	A	C	G	C/T	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	4
5	G	C	-	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	1	5
6	G	C	G	G	G	A	C	G	C	C/T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	6
7	G	C	G/C	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	1	7
8	G	C	G	G	G	A	C	G/A	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	8
9	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G/T	1	9
10	G	C	G	G	G	A	C	G	C	C/T	T	C	C	C	A	C	G	G	G	C	C	G/A	G	G	T	A	G	1	10
11	G	C	G	G	G	A/G	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T/G	A	G	1	11
12	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C/T	G	G	G	C	C	G	G	G	T	A/T	G	1	12
13	G	C	G/C	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G/T	1	13
14	G	C	G/C	G	G	A	C	G	C	C	T	C	C	C	A	C	G/T	G	G	C	C	G	G	G	T	A	G	1	14
15	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A/G	C/T	G	G	G	C	C	G	G	G	T	A/T	G	1	15
16	G	C	G	G	G	A/G	C	G	C	C	T	C	C	C	A	C/T	G	G	G	C	C	G	G	G	T	A/T	G	1	16
17	G	C	G	G	G	A	C/T	G/A	C	C	T	C/T	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	17
18	G	C	G	G	G	A/G	C	G	C/T	C	T/C	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	18
19	G	C/T	G	G	G	A/G	C/T	G	C	C	T/C	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	19
20	G/A	C	G	G	G	A/G	C	G	C	C	T/C	C	C	C	A	C	G	G	G	C	C/T	G	G	G	T	A	G	1	20
21	G/A	C	G	G	G	A	C	G	C/T	C	T	C	C	C	A	C	G	G	G	C/T	C/T	G	G	G	T	A	G	1	21
22	G/A	C	G	G	G	A	C	G	C	C/T	T	C	C	C	A	C	G	G	G	C	C/T	G	G	G	T	A	G	1	22
23	G	C	G/C	G	G	A/G	C	G	C	C	T/C	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	1	23
24	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G/T	T	A/T	G	1	24
25	G	C	G	G	G	A/G	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G/T	1	25
26	G	C	G	-	G	A	C	G	C	-	-	-	C	C/A	A	C	G	G	G	C	C	G	G	G	T	A/T	G	1	26
27	G	C	G	G	G	A	C	G/A	C	C	T	C	C	C	A	C/T	G	G	G	C	C	G	G	G	T	A/T	G	1	27
28	G	C	G	G/A	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	28
29	G	C	G	G	G	A	C	G	C/T	C	T	C	C	C	A	C/T	G	G	G	C	C	G	G	G	T	A/T	G	1	29
30	G	C	G	G	G	A	C/T	G	C	C	T	C	C	C	A	C/T	G	G	G	C	C	G	G	G	T	A/T	G	1	30
31	G	C	G	G	G	A	C	G/A	C	C/T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	1	31
32	G	-	G	G	G	A	C	G	C/T	C	T	C	C	C	A	C	G	G	G/A	C	C	G	G	G	T	A	G	1	32
33	G	C	G	-	G	A	C	G	C	C	T	C	C	C	A	C	G/T	G	-	C	C	G	G	G	T	A	G	1	33
34	G	C	G	G	G/A	A/G	C	G	C	C	T	-	C	C	A/G	C	G	G	G	C	C	G	G	G	T	A	G	1	34
35	G	C	G	G	G	A	C	G	C	C/T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	1	35
36	G	C	G	G	G	G	C	G	C	C	T	C	C	C	G	C	G	G	G	C	C	G	G	G	T	A	G	2	2
37	G	C	G	G	G	A/G	C/T	G	C	C	T	C	C	C	A/G	C	G	G	G	C	C	G	G	G	T	A	G	2	3

Table 3 (cont'd). Genotypes and Haplotype Pairs Observed in the APOA4 Gene																													
Genotype	Polymorphic Sites																											HAP	
Number	PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16	PS 17	PS 18	PS 19	PS 20	PS 21	PS 22	PS 23	PS 24	PS 25	PS 26	PS 27	PAIR	
38	G	C	G	G	G	A/G	C	G	C/T	C	T	C	C	C	A/G	C	G	G	G	C	C	G	G	G	T	A	G	2	4
39	G	C	G	G	G	A/G	C	G	C	C	T	C	C	C	A/G	C	G	G	-	C	C	G	G	G	T	A/T	G	2	5
40	G	C	G	G	G	A/G	C	G	C	C/T	T	C	C	C	A/G	C	G	G	G	C	C	G	G	G	T	A	G	2	6
41	G	C	G	G	G	A/G	C	G/A	-	C	T	C	C	C	A/G	C	G	G	G	C	C	G	G	G	T	A	G	2	8
42	G	C	G	-	G	G	C	G	C	C	T	C	C/T	C	G	C	G	G/A	G	C	C	G	G	G	T	A	G	2	36
43	G	C	G	G	G	A	T	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	3	3
44	G	C	G	-	G	A	C/T	G	C/T	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	3	4
45	G	C	G	G	G	A	C/T	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	3	5
46	G	C	G/C	G	G	A	C/T	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	3	7
47	G	C	G	G	G	A	C	G	T	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	4	4
48	G	C	G	G	G	A	C	G	C/T	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G	4	5
49	G	C	G	G	G	A	C	G	C/T	C/T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A/T	G	4	6
50	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G/A	G	T	A	G	5	5
51	G	C	G/C	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	T	G	5	7

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family) and then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 36 human APOA4 haplotypes shown in Table 4 below.

Table 4. Haplotypes Observed in the APOA4 Gene

Haplotype	Polymorphic Sites																										
Number	PS 1	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 16	PS 17	PS 18	PS 19	PS 20	PS 21	PS 22	PS 23	PS 24	PS 25	PS 26	PS 27
1	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
2	G	C	G	G	G	G	C	G	C	C	T	C	C	C	G	C	G	G	G	C	C	G	G	G	T	A	G
3	G	C	G	G	G	A	T	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
4	G	C	G	G	G	A	C	G	T	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
5	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	T	G
6	G	C	G	G	G	A	C	G	C	T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
7	G	C	C	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	T	G
8	G	C	G	G	G	A	C	A	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
9	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	T
10	G	C	G	G	G	A	C	G	C	T	T	C	C	C	A	C	G	G	G	C	C	A	G	G	T	A	G
11	G	C	G	G	G	G	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	G	A	G
12	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	T	G	G	G	C	C	G	G	G	T	T	G
13	G	C	C	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	T	T
14	G	C	C	G	G	A	C	G	C	C	T	C	C	C	A	C	T	G	G	C	C	G	G	G	T	A	G
15	G	C	G	G	G	A	C	G	C	C	T	C	C	C	G	T	G	G	G	C	C	G	G	G	T	T	G
16	G	C	G	G	G	G	C	G	C	C	T	C	C	C	A	T	G	G	G	C	C	G	G	G	T	T	G
17	G	C	G	G	G	A	T	A	C	C	T	T	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
18	G	C	G	G	G	G	C	G	T	C	C	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
19	G	T	G	G	G	G	T	G	C	C	C	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
20	A	C	G	G	G	G	C	G	C	C	C	C	C	C	A	C	G	G	G	C	T	G	G	G	T	A	G
21	A	C	G	G	G	A	C	G	T	C	T	C	C	C	A	C	G	G	G	T	T	G	G	G	T	A	G
22	A	C	G	G	G	A	C	G	C	T	T	C	C	C	A	C	G	G	G	C	T	G	G	G	T	A	G
23	G	C	C	G	G	G	C	G	C	C	C	C	C	C	A	C	G	G	G	C	C	G	G	G	T	T	G
24	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	T	T	T	G
25	G	C	G	G	G	G	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	T
26	G	C	G	G	G	A	C	G	C	C	T	C	C	A	A	C	G	G	G	C	C	G	G	G	T	T	G
27	G	C	G	G	G	A	C	A	C	C	T	C	C	C	A	T	G	G	G	C	C	G	G	G	T	T	G
28	G	C	G	A	G	A	C	G	C	C	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
29	G	C	G	G	G	A	C	G	T	C	T	C	C	C	A	T	G	G	G	C	C	G	G	G	T	T	G
30	G	C	G	G	G	A	T	G	C	C	T	C	C	C	A	T	G	G	G	C	C	G	G	G	T	T	G
31	G	C	G	G	G	A	C	A	C	T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	A	G
32	G	C	G	G	G	A	C	G	T	C	T	C	C	C	A	C	G	G	A	C	C	G	G	G	T	A	G
33	G	C	G	G	G	A	C	G	C	C	T	C	C	C	A	C	T	G	G	C	C	G	G	G	T	A	G
34	G	C	G	G	A	G	C	G	C	C	T	C	C	C	G	C	G	G	G	C	C	G	G	G	T	A	G
35	G	C	G	G	G	A	C	G	C	T	T	C	C	C	A	C	G	G	G	C	C	G	G	G	T	T	G
36	G	C	G	G	G	G	C	G	C	C	T	C	T	C	G	C	G	A	G	C	C	G	G	G	T	A	G

Table 5 below shows the number of chromosomes seen for each haplotype arranged by the ethnic background of the subjects in the Index Repository.

Table 5. Frequencies of Observed Haplotypes in the APOA4 Gene						
Haplotype Number	AF	AS	CA	HL	AM	Total
1	17	14	19	19	4	75
2	1	10	6	4	0	21
3	0	9	0	2	0	11
4	3	3	0	2	0	8
5	1	0	6	1	0	8
6	4	0	0	0	0	4
7	1	0	1	2	0	4
8	1	0	2	0	0	3
9	0	0	2	1	0	3
10	2	0	0	0	0	2
11	2	0	0	0	0	2
12	1	0	0	0	0	1
13	0	0	1	0	0	1
14	0	0	1	0	0	1
15	0	0	0	1	0	1
16	0	0	0	0	1	1
17	0	1	0	0	0	1
18	0	0	1	0	0	1
19	0	1	0	0	0	1
20	0	0	0	1	0	1
21	1	0	0	0	0	1
22	1	0	0	0	0	1
23	0	0	1	0	0	1
24	1	0	0	0	0	1
25	1	0	0	0	0	1
26	1	0	0	0	0	1
27	0	0	1	0	0	1
28	0	0	0	0	1	1
29	0	0	1	0	0	1
30	0	0	0	1	0	1
31	1	0	0	0	0	1
32	0	1	0	0	0	1
33	1	0	0	0	0	1
34	0	0	0	1	0	1
35	0	0	0	1	0	1
36	0	1	0	0	0	1

Table 6 below shows the number of subjects seen for a given haplotype pair arranged by the ethnic background of the subjects in the Index Repository.

Table 6. Frequencies of Haplotype Pairs Observed in the APOA4 Gene							
HAP PAIR		AF	AS	CA	HL	AM	Total
1	1	0	2	2	5	1	10
2	1	0	3	6	1	0	10
2	2	0	2	0	1	0	3
3	1	0	2	0	0	0	2
3	2	0	2	0	0	0	2
3	3	0	2	0	0	0	2
4	1	3	2	0	1	0	6
4	2	0	0	0	1	0	1
4	3	0	1	0	0	0	1
4	4	1	0	0	0	0	1
5	1	0	0	1	0	0	1
5	2	0	0	2	0	0	2
5	3	0	0	0	1	0	1
5	4	1	0	0	0	0	1
5	5	0	0	1	0	0	1
6	1	2	0	0	0	0	2
6	2	1	0	0	0	0	1
6	4	3	0	0	0	0	3
7	1	1	0	0	1	0	2
7	3	0	0	0	1	0	1
7	5	0	0	1	0	0	1
8	1	1	0	5	0	0	6
8	2	0	0	1	0	0	1
9	1	0	0	2	1	0	3
10	1	2	0	0	0	0	2
11	1	2	0	0	0	0	2
12	1	1	0	0	0	0	1
13	1	0	0	1	0	0	1
14	1	0	0	1	0	0	1
15	1	0	0	0	1	0	1
16	1	0	0	0	0	1	1
17	1	0	1	0	0	0	1
18	1	0	0	1	0	0	1
19	1	0	1	0	0	0	1
20	1	0	0	0	1	0	1
21	1	1	0	0	0	0	1
22	1	1	0	0	0	0	1
23	1	0	0	1	0	0	1
24	1	1	0	0	0	0	1
25	1	1	0	0	0	0	1
26	1	1	0	0	0	0	1
27	1	0	0	1	0	0	1
28	1	0	0	0	0	1	1
29	1	0	0	1	0	0	1
30	1	0	0	0	1	0	1
31	1	1	0	0	0	0	1
32	1	0	1	0	0	0	1
33	1	1	0	0	0	0	1
34	1	0	0	0	1	0	1
35	1	0	0	0	1	0	1
36	2	0	1	0	0	0	1

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the apolipoprotein A-IV (APOA4) gene of an individual which comprises determining whether the individual has one of the APOA4 haplotypes shown in Table 4 or one of the haplotype pairs shown in Table 3.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-27 on at least one copy of the individual's APOA4 gene.

3. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-27 on both copies of the individual's APOA4 gene.

4. A method for genotyping the apolipoprotein A-IV (APOA4) gene of an individual, comprising determining for the two copies of the APOA4 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25.

5. The method of claim 4, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the APOA4 gene, or a fragment thereof, that are present in the individual;
   (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
   (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

6. The method of claim 4, which comprises determining for the two copies of the APOA4 gene present in the individual the identity of the nucleotide pair at each of PS1-27.

7. A method for haplotyping the apolipoprotein A-IV (APOA4) gene of an individual which comprises determining, for one copy of the APOA4 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25.

8. The method of claim 7, further comprising determining the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4, PS6, PS7, PS9, PS10, PS11, PS15, PS16, PS17, PS20, PS21, PS22, PS24, PS26, and PS27.

9. The method of claim 7, wherein the determining step comprises:

(a) isolating from the individual a nucleic acid sample containing only one of the two copies of the APOA4 gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid molecule a target region containing the selected polymorphic site;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

10. A method for predicting a haplotype pair for the apolipoprotein A-IV (APOA4) gene of an individual comprising:

(a) identifying an APOA4 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the data in Table 3; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-27.

12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the apolipoprotein A-IV (APOA4) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-36 shown in Table 4 and the haplotype pair is selected from the haplotype pairs shown in Table 3, wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

13. The method of claim 12, wherein the trait is a clinical response to a drug targeting APOA4.

14. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the apolipoprotein A-IV (APOA4) gene at a polymorphic site selected from the group consisting of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25.

15. The composition of claim 14, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the APOA4 gene at a region containing the polymorphic site.

16. The composition of claim 15, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-13, the complements of SEQ ID.

NOS:4-13, and SEQ ID NOS:14-33.

17. The composition of claim 14, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.

18. The composition of claim 17, wherein the primer extension oligonucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:34-53.

19. A kit for genotyping the APOA4 gene of an individual, which comprises a set of oligonucleotides designed to genotype each of PS2, PS5, PS8, PS12, PS13, PS14, PS18, PS19, PS23, and PS25.

20. The kit of claim 19, which further comprises oligonucleotides designed to genotype each of PS1, PS3, PS4, PS6, PS7, PS9, PS10, PS11, PS15, PS16, PS17, PS20, PS21, PS22, PS24, PS26, and PS27.

21. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for the apolipoprotein A-IV (APOA4) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises an APOA4 isogene defined by a haplotype selected from the group consisting of haplotypes 1-36 in Table 4; and

a second nucleotide sequence which is complementary to the first nucleotide sequence.

22. The isolated polynucleotide of claim 21, which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

23. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 21, wherein the organism expresses an APOA4 protein encoded by the first nucleotide sequence.

24. The recombinant organism of claim 23, which is a nonhuman transgenic animal.

25. The isolated polynucleotide of claim 21, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the APOA4 gene, the fragment comprising one or more polymorphisms selected from the group consisting of thymine at PS2, adenine at PS5, adenine at PS8, thymine at PS12, thymine at PS13, adenine at PS14, adenine at PS18, adenine at PS19, adenine at PS23, and guanine at PS25.

26. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the APOA4 cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises the coding sequence of an APOA4 isogene defined by one of the haplotypes shown in Table 4.

27. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide

of claim 26, wherein the organism expresses a apolipoprotein A-IV (APOA4) protein encoded by the polymorphic variant sequence.

28. The recombinant organism of claim 27, which is a nonhuman transgenic animal.

29. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the APOA4 protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant is encoded by an isogene defined by one of the haplotypes shown in Table 4.

30. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 29.

31. A method for screening for drugs targeting the isolated polypeptide of claim 29 which comprises contacting the APOA4 polymorphic variant with a candidate agent and assaying for binding activity.

32. A computer system for storing and analyzing polymorphism data for the apolipoprotein A-IV gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 4.

33. A genome anthology for the apolipoprotein A-IV (APOA4) gene which comprises APOA4 isogenes defined by any one of haplotypes 1-36 shown in Table 4.

POLYMORPHISMS IN THE APOA4 GENE

```
TCTAGAGATT TAAAGTATGT GGGAGGCTGT GCTTAGGTTA TATGTAAATA
CTATGCCATT TTATATCAAG GACTTGAACA TCCATGGATT TTGGTATCTG   100
CAGAGGGTCC TGGAATCATT TCCCATGGAG ACTGAGAGAT GACCGTACTA
CCCACTTCGC AAGCAATGTC TTCTTTAATG TACTGAACCA TCCCATTGTT   200
CAGAGGAGAA ACTGAAGCTC AGGGCTTTGA ATAACTAGAC CAAGGAGGCA
CAGCATGGGA GTGGGACATG AAGCACTCTA CAATTAACCC TTTCAGGACA   300
AGGCCCTGTC TCCCACACCC ATCTGCCCAA AGGCTCTCCA GGGCCCCCTC
CTCTTGGGTG TACCTTGACA AGAGACCTAG ATTTTAGCTC ACTATGCTGT   400
CTGCAGTCCT GGATGGTCCC ACTCCAGTGT CTGGTGCTCT GAGATGGAGT
CAGCATTAGT GGCGGATGTG GAGACTGGGG GGACCTGTCT TCACTGGGGT   500
AGACAGAGGA GATGTGGACT TTGCCCCCCA TGAGCCCGGC ACAAACCCGA
GAGGCCGCCA GCAGGGCCTC GAGGCATCAG TCCCGGGCTC ATGGGCTCCC   600
                      A
TGAGGTGTTT CTCCTACTGT TTTCCGTTCC CCTCCTCCCT TCCATGCTGA
GGTTGGTGGG GTGGGGGTGG GGGTGCCCAC GCACGGAACA GCCACCACTT   700
                                                   T
CTAACCTATC GCCTGAGCCC TGATCTGCTG TCAGCTTCCA CGTAGTCTCA
GGGTCACAAA AGTCCAAGAG GCCTCTTGGG AATGTGTCAC CTTCCAGCGT   800
GGAGTCACAC TGAGGAAGGA GGAGGGGAGG GCAGCCAGGG GGGTGGCGAT
AGGGAGAGAG TTTAAATGTC TGGCTGGCTC TGAGCTTCAG TCAGTTCCCA   900
    C
CTGCAGCGCA GGTGAGCTCT CCTGAGGACC TCTCTGTCAG CTCCCCTGAT
    A
TGTAGGGAGG ATCCAGTGTG GCAAGAAACT CCTCCAGCCC AGCAAGCAGC   1000
TCAGGATGTT CCTGAAGGCC GTGGTCCTGA CCCTGGCCCT GGTGGCTGTC
                                              A
    [exon 1: 1006..
GCCGGTGAGT AGAAGCTGTC TTTGGATGGC ACTCCTGGGC TGCTGCTCTG   1100
             ..1054]
AGTAGTGCAG GATGGAGGCT GAGCCAAAGC AAAAGGACAC TTCTGAGTGC
CCATCAGCCC CCAGCTGGAC ATGAGGTCTG CCTGGCTGCC AAGTGGCTCA   1200
CAGGAGAGCT GGCCCAGTCC CAGTGGTGGG CCCATTGGCA TTGGTGCTAT
ACCAGTTTCA CATATCCCTG TGGCTTCCAA AAAGCTAAGC TCAGACAGGG   1300
AAAATGGCAG GTTGAGGCAC CCCCACCATC ATCCAGTCTG CAGCTCAGAG
CTGGAGCAGA GGGGCCACAC AGGAGACGGG GCCTCATGAA TTGCTCTCTG   1400
TTACCACCCA GGAGCCAGGG CTGAGGTCAG TGCTGACCAG GTGGCCACAG
                                                  G
    [exon 2: 1412..
TGATGTGGGA CTACTTCAGC CAGCTGAGCA ACAATGCCAA GGAGGCCGTG   1500
GAACATCTCC AGAAATCTGA ACTCACCCAG CAACTCAAGT AAGAGGGACT
             ..1538]
ACAGTGTGTG GTGGTGACGG GGAATTCTTA AAGGCCATGC AATGTACTGG   1600
      CA
CAAGGGTTGA GCTTAGAGAC AGGAGCCCTG AGCTTAGGAT ACCCACTGCC
CTGCCACTAA CTGGCCGGGC CTCTGAACCT AGGATCCACA TATGTAAACC   1700
                                                    T
GGAAGTTTGG ACCGAATAAT CCCTGCCATG TCCTTTTGCT TTGACGTTCT
AGAGTTTGAC AAATGGCCAC ATCCTATCAT TCAGGCTCAT GGAAGAGAGG   1800
GAGGGAGGAA AATGTCATGT GAGCTGATTT CTAATACGTT TCAGAAAGAC
AGGCCCCAGT GGAATCAAGG GGAGGGAGGT GGGAATATTT GGGAGGCCCC   1900
TGGGCACAGG CAAGGAAAGC AGCACCTTGT GCCACTGGAA GACCCCAGCA
GAGGTCAAGA AGACAACATT GTGTTACACA ATGTGATCCT ATGGCCCAGA   2000
```

FIGURE 1A

```
ACATTCCCTC TGGGAAGGAC CTCAAAGTCC CACCCTCTGC AGACAAGGAG
GGGAAAGCAA ACTGCTGGAG GTGACATGGT GGGTAGATTC TGAGACAAAC   2100
TATGTGGGAG ATCCTGAGAT AGAAATTCAG CATCGTAACT TAGTCTGTGA
CACCCATCCT CTCCAATCTG CACCACCATA GGGAGGGTGA ACTCGGTACC   2200
TCTGAGCACT CACCTGTCTT AGCACGTGTG CATAAGGCGA GTGGTATACA
AGCAGACAAA GTCTTGCCGT GTAAATGCCA AATGTAACGT GGCCTCCTTG   2300
TGCCCTTCCC CACAGTGCCC TCTTCCAGGA CAAACTTGGA GAAGTGAACA
    [exon 3: 2316..
CTTACGCAGG TGACCTGCAG AAGAAGCTGG TGCCCTTTGC CACCGAGCTG   2400
    T      C
CATGAACGCC TGGCCAAGGA CTCGGAGAAA CTGAAGGAGG AGATTGGGAA
GGAGCTGGAG GAGCTGAGGG CCCGGCTGCT GCCCCATGCC AATGAGGTGA   2500
              T       T
GCCAGAAGAT CGGGGACAAC CTGCGAGAGC TTCAGCAGCG CCTGGAGCCC
                    A
TACGCGGACC AGCTGCGCAC CCAGGTCAAC ACGCAGGCCG AGCAGCTGCG   2600
                              G          T
GCGCCAGCTG ACCCCCTACG CACAGCGCAT GGAGAGAGTG CTGCGGGAGA
                    T
ACGCCGACAG CCTGCAGGCC TCGCTGAGGC CCCACGCCGA CGAGCTCAAG   2700
GCCAAGATCG ACCAGAACGT GGAGGAGCTC AAGGGACGCC TTACGCCCTA
                   A  A
CGCTGACGAA TTCAAAGTCA AGATTGACCA GACCGTGGAG GAGCTGCGCC   2800
GCAGCCTGGC TCCCTATGCT CAGGACACGC AGGAGAAGCT CAACCACCAG
CTTGAGGGCC TGACCTTCCA GATGAAGAAG AACGCCGAGG AGCTCAAGGC   2900
CAGGATCTCG GCCAGTGCCG AGGAGCTGCG GCAGAGGCTG GCGCCCTTGG
                   T
CCGAGGACGT GCGTGGCAAC CTGAGGGGCA ACACCGAGGG GCTGCAGAAG   3000
                                     T
TCACTGGCAG AGCTGGGTGG GCACCTGGAC CAGCAGGTGG AGGAGTTCCG
                            A         A
ACGCCGGGTG GAGCCCTACG GGGAAAACTT CAACAAAGCC CTGGTGCAGC   3100
AGATGGAACA GCTCAGGACG AAACTGGGCC CCCATGCGGG GGACGTGGAA
                                                 T
GGCCACTTGA GCTTCCTGGA GAAGGACCTG AGGGACAAGG TCAACTCCTT   3200
                                                 G
CTTCAGCACC TTCAAGGAGA AAGAGAGCCA GGACAAGACT CTCTCCCTCC
                                        T
CTGAGCTGGA GCAACAGCAG GAACAGCATC AGGAGCAGCA GCAGGAGCAG   3300
                              G
GTGCAGATGC TGGCCCCTTT GGAGAGCTGA GCTGCCCCTG GTGCACTGGC
               ..3330].
CCCACCCTCG TGGACACCTG CCCTGCCCTG CCACCTGTCT GTCTGTCTGT   3400
CCCAAAGAAG TTCTGGTATG AACTTGAGGA CACATGTCCA GTGGGAGGTG
AGACCACCTC TCAATATTCA ATAAAGCTGC TGAGAATCTA GCCTCAACTG   3500
GTTGCCGGAT GAATCCTCCT TGCAGCTGGG GAGGTGGGGA GGTAACCATG
ACTGGGCAGA GCTTAGCAGC GGCCTGGCAG GAGACACCCA GGATTGGGGA   3600
GATGAGACTG CAG                                           3613
```

FIGURE 1B

POLYMORPHISMS IN THE CODING SEQUENCE OF APOA4

```
ATGTTCCTGA AGGCCGTGGT CCTGACCCTG GCCCTGGTGG CTGTCGCCGG
                                       A
AGCCAGGGCT GAGGTCAGTG CTGACCAGGT GGCCACAGTG ATGTGGGACT      100
                                       G
ACTTCAGCCA GCTGAGCAAC AATGCCAAGG AGGCCGTGGA ACATCTCCAG
AAATCTGAAC TCACCCAGCA ACTCAATGCC CTCTTCCAGG ACAAACTTGG      200
AGAAGTGAAC ACTTACGCAG GTGACCTGCA GAAGAAGCTG GTGCCCTTTG
                T      C
CCACCGAGCT GCATGAACGC CTGGCCAAGG ACTCGGAGAA ACTGAAGGAG      300
GAGATTGGGA AGGAGCTGGA GGAGCTGAGG GCCCGGCTGC TGCCCCATGC
                          T       T
CAATGAGGTG AGCCAGAAGA TCGGGGACAA CCTGCGAGAG CTTCAGCAGC      400
                                 A
GCCTGGAGCC CTACGCGGAC CAGCTGCGCA CCCAGGTCAA CACGCAGGCC
                                          G          T
GAGCAGCTGC GGCGCCAGCT GACCCCCTAC GCACAGCGCA TGGAGAGAGT      500
                                 T
GCTGCGGGAG AACGCCGACA GCCTGCAGGC CTCGCTGAGG CCCCACGCCG
ACGAGCTCAA GGCCAAGATC GACCAGAACG TGGAGGAGCT CAAGGGACGC      600
                               A  A
CTTACGCCCT ACGCTGACGA ATTCAAAGTC AAGATTGACC AGACCGTGGA
GGAGCTGCGC CGCAGCCTGG CTCCCTATGC TCAGGACACG CAGGAGAAGC      700
TCAACCACCA GCTTGAGGGC CTGACCTTCC AGATGAAGAA GAACGCCGAG
GAGCTCAAGG CCAGGATCTC GGCCAGTGCC GAGGAGCTGC GGCAGAGGCT      800
                               T
GGCGCCCTTG GCCGAGGACG TGCGTGGCAA CCTGAGGGGC AACACCGAGG
                                                 T
GGCTGCAGAA GTCACTGGCA GAGCTGGGTG GGCACCTGGA CCAGCAGGTG      900
                                        A         A
GAGGAGTTCC GACGCCGGGT GGAGCCCTAC GGGGAAAACT TCAACAAAGC
CCTGGTGCAG CAGATGGAAC AGCTCAGGAC GAAACTGGGC CCCCATGCGG      1000
GGGACGTGGA AGGCCACTTG AGCTTCCTGG AGAAGGACCT GAGGGACAAG
       T
GTCAACTCCT TCTTCAGCAC CTTCAAGGAG AAAGAGAGCC AGGACAAGAC      1100
      G                                             T
TCTCTCCCTC CCTGAGCTGG AGCAACAGCA GGAACAGCAT CAGGAGCAGC
                                          G
AGCAGGAGCA GGTGCAGATG CTGGCCCCTT TGGAGAGCTG A               1191
```

FIGURE 2

ISOFORMS OF THE APOA4 PROTEIN

```
MFLKAVVLTL ALVAVAGARA EVSADQVATV MWDYFSQLSN NAKEAVEHLQ
             M
KSELTQQLNA LFQDKLGEVN TYAGDLQKKL VPFATELHER LAKDSEKLKE      100
EIGKELEELR ARLLPHANEV SQKIGDNLRE LQQRLEPYAD QLRTQVNTQA
           V                K                     S
EQLRRQLTPY AQRMERVLRE NADSLQASLR PHADELKAKI DQNVEELKGR      200
           S                                   X
LTPYADEFKV KIDQTVEELR RSLAPYAQDT QEKLNHQLEG LTFQMKKNAE
ELKARISASA EELRQRLAPL AEDVRGNLRG NTEGLQKSLA ELGGHLDQQV      300
EEFRRRVEPY GENFNKALVQ QMEQLRTKLG PHAGDVEGHL SFLEKDLRDK
VNSFFSTFKE KESQDKTLSL PELEQQQEQH QEQQQEQVQM LAPLES          396
  A              S             Q
```

X= M or I

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Bentivegna, Steven C.
Choi, Julie Y.
Kliem, Stefanie E.
Koshy, Beena

<120> HAPLOTYPES OF THE APOA4 GENE

<130> MWH-0506PCT APOA4

<140> TBA
<141> 2001-04-03

<150> 60/194,362
<151> 2000-04-05

<160> 56

<170> PatentIn Ver. 2.1

<210> 1
<211> 3613
<212> DNA
<213> Homo sapien

<400> 1
```
tctagagatt taaagtatgt gggaggctgt gcttaggtta tatgtaaata ctatgccatt 60
ttatatcaag gacttgaaca tccatggatt ttggtatctg cagagggtcc tggaatcatt 120
tcccatggag actgagagat gaccgtacta cccacttcgc aagcaatgtc ttctttaatg 180
tactgaacca tcccattgtt cagaggagaa actgaagctc agggctttga ataactagac 240
caaggaggca cagcatggga gtgggacatg aagcactcta caattaaccc tttcaggaca 300
aggccctgtc tcccacaccc atctgcccaa aggctctcca gggccccctc ctcttgggtg 360
taccttgaca agagacctag attttagctc actatgctgt ctgcagtcct ggatggtccc 420
actccagtgt ctggtgctct gagatggagt cagcattagt ggcggatgtg gagactgggg 480
ggacctgtct tcactggggt agacagagga gatgtggact ttgcccccca tgagcccggc 540
acaaacccga gaggccgcca gcagggcctc gaggcatcag tcccgggctc atgggctccc 600
tgaggtgttt ctcctactgt tttccgttcc cctcctccct tccatgctga ggttggtggg 660
gtgggggtgg gggtgcccac gcacggaaca gccaccactt ctaacctatc gcctgagccc 720
tgatctgctg tcagcttcca cgtagtctca gggtcacaaa agtccaagag gcctcttggg 780
aatgtgtcac cttccagcgt ggagtcacac tgaggaagga ggaggggagg gcagccaggg 840
gggtggcgat agggagagag tttaaatgtc tggctggctc tgagcttcag tcagttccca 900
ctgcagcgca ggtgagctct cctgaggacc tctctgtcag ctcccctgat tgtagggagg 960
atccagtgtg gcaagaaact cctccagccc agcaagcagc tcaggatgtt cctgaaggcc 1020
gtggtcctga ccctggccct ggtggctgtc gccggtgagt agaagctgtc tttggatggc 1080
actcctgggc tgctgctctg agtagtgcag gatggaggct gagccaaagc aaaaggacac 1140
ttctgagtgc ccatcagccc ccagctggac atgaggtctg cctggctgcc aagtggctca 1200
```

caggagagct ggcccagtcc cagtggtggg cccattggca ttggtgctat accagtttca 1260
catatccctg tggcttccaa aaagctaagc tcagacaggg aaaatggcag gttgaggcac 1320
ccccaccatc atccagtctg cagctcagag ctggagcaga ggggccacac aggagacggg 1380
gcctcatgaa ttgctctctg ttaccaccca ggagccaggg ctgaggtcag tgctgaccag 1440
gtggccacag tgatgtggga ctacttcagc cagctgagca acaatgccaa ggaggccgtg 1500
gaacatctcc agaaatctga actcacccag caactcaagt aagagggact acagtgtgtg 1560
gtggtgacgg ggaattctta aaggccatgc aatgtactgg caagggttga gcttagagac 1620
aggagccctg agcttaggat acccactgcc ctgccactaa ctggccgggc ctctgaacct 1680
aggatccaca tatgtaaacc ggaagtttgg accgaataat ccctgccatg tccttttgct 1740
ttgacgttct agagtttgac aaatggccac atcctatcat tcaggctcat ggaagagagg 1800
gagggaggaa aatgtcatgt gagctgattt ctaatacgtt tcagaaagac aggccccagt 1860
ggaatcaagg ggagggaggt gggaatattt gggaggcccc tgggcacagg caaggaaagc 1920
agcaccttgt gccactggaa gaccccagca gaggtcaaga agacaacatt gtgttacaca 1980
atgtgatcct atggcccaga acattccctc tgggaaggac ctcaaagtcc cacccctctgc 2040
agacaaggag gggaaagcaa actgctggag gtgacatggt gggtagattc tgagacaaac 2100
tatgtgggag atcctgagat agaaattcag catcgtaact tagtctgtga cacccatcct 2160
ctccaatctg caccaccata gggagggtga actcggtacc tctgagcact cacctgtctt 2220
agcacgtgtg cataaggcga gtggtataca agcagacaaa gtcttgccgt gtaaatgcca 2280
aatgtaacgt ggcctccttg tgcccttccc cacagtgccc tcttccagga caaacttgga 2340
gaagtgaaca cttacgcagg tgacctgcag aagaagctgg tgccctttgc caccgagctg 2400
catgaacgcc tggccaagga ctcggagaaa ctgaaggagg agattgggaa ggagctggag 2460
gagctgaggg cccggctgct gccccatgcc aatgaggtga gccagaagat cggggacaac 2520
ctgcgagagc ttcagcagcg cctggagccc tacgcggacc agctgcgcac ccaggtcaac 2580
acgcaggccg agcagctgcg gcgccagctg accccctacg cacagcgcat ggagagagtg 2640
ctgcgggaga acgccgacag cctgcaggcc tcgctgaggc cccacgccga cgagctcaag 2700
gccaagatcg accagaacgt ggaggagctc aagggacgcc ttacgcccta cgctgacgaa 2760
ttcaaagtca agattgacca gaccgtggag gagctgcgcc gcagcctggc tccctatgct 2820
caggacacgc aggagaagct caaccaccag cttgagggcc tgaccttcca gatgaagaag 2880
aacgccgagg agctcaaggc caggatctcg gccagtgccg aggagctgcg gcagaggctg 2940
gcgccccttgg ccgaggacgt gcgtggcaac ctgaggggca acaccgaggg gctgcagaag 3000
tcactggcag agctgggtgg gcacctggac cagcaggtgg aggagttccg acgccgggtg 3060
gagccctacg gggaaaactt caacaaagcc ctggtgcagc agatggaaca gctcaggacg 3120
aaactgggcc cccatgcggg ggacgtggaa ggccacttga gcttcctgga gaaggacctg 3180
agggacaagg tcaactcctt cttcagcacc ttcaaggaga aagagagcca ggacaagact 3240
ctctccctcc ctgagctgga gcaacagcag gaacagcatc aggagcagca gcaggagcag 3300
gtgcagatgc tggccccttt ggagagctga gctgcccctg gtgcactggc cccaccctcg 3360
tggacacctg ccctgccctg ccacctgtct gtctgtctgt cccaaagaag ttctggtatg 3420
aacttgagga cacatgtcca gtgggaggtg agaccacctc tcaatattca ataaagctgc 3480
tgagaatcta gcctcaactg gttgccggat gaatcctcct tgcagctggg gaggtgggga 3540
ggtaaccatg actgggcaga gcttagcagc ggcctggcag gagacaccca ggattgggga 3600
gatgagactg cag 3613

<210> 2
<211> 1191
<212> DNA
<213> Homo sapien

<400> 2
atgttcctga aggccgtggt cctgaccctg gccctggtgg ctgtcgccgg agccagggct 60
gaggtcagtg ctgaccaggt ggccacagtg atgtgggact acttcagcca gctgagcaac 120
aatgccaagg aggccgtgga acatctccag aaatctgaac tcacccagca actcaatgcc 180
ctcttccagg acaaacttgg agaagtgaac acttacgcag gtgacctgca gaagaagctg 240
gtgccctttg ccaccgagct gcatgaacgc ctggccaagg actcggagaa actgaaggag 300
gagattggga aggagctgga ggagctgagg gcccggctgc tgccccatgc caatgaggtg 360
agccagaaga tcggggacaa cctgcgagag cttcagcagc gcctggagcc ctacgcggac 420
cagctgcgca cccaggtcaa cacgcaggcc gagcagctgc ggcgccagct gacccccctac 480
gcacagcgca tggagagagt gctgcgggag aacgccgaca gcctgcaggc ctcgctgagg 540
ccccacgccg acgagctcaa ggccaagatc gaccagaacg tggaggagct caagggacgc 600
cttacgccct acgctgacga attcaaagtc aagattgacc agaccgtgga ggagctgcgc 660
cgcagcctgg ctccctatgc tcaggacacg caggagaagc tcaaccacca gcttgagggc 720
ctgaccttcc agatgaagaa gaacgccgag gagctcaagg ccaggatctc ggccagtgcc 780
gaggagctgc ggcagaggct ggcgcccttg gccgaggacg tgcgtggcaa cctgaggggc 840
aacaccgagg ggctgcagaa gtcactggca gagctgggtg ggcacctgga ccagcaggtg 900
gaggagttcc gacgccgggt ggagccctac ggggaaaact tcaacaaagc cctggtgcag 960
cagatggaac agctcaggac gaaactgggc ccccatgcgg gggacgtgga aggccacttg 1020
agcttcctgg agaaggacct gagggacaag gtcaactcct tcttcagcac cttcaaggag 1080
aaagagagcc aggacaagac tctctccctc cctgagctgg agcaacagca ggaacagcat 1140
caggagcagc agcaggagca ggtgcagatg ctggcccctt tggagagctg a 1191

<210> 3
<211> 396
<212> PRT
<213> Homo sapien

<400> 3
Met Phe Leu Lys Ala Val Val Leu Thr Leu Ala Leu Val Ala Val Ala
1 5 10 15

Gly Ala Arg Ala Glu Val Ser Ala Asp Gln Val Ala Thr Val Met Trp
20 25 30

Asp Tyr Phe Ser Gln Leu Ser Asn Asn Ala Lys Glu Ala Val Glu His
35 40 45

Leu Gln Lys Ser Glu Leu Thr Gln Gln Leu Asn Ala Leu Phe Gln Asp
50 55 60

Lys Leu Gly Glu Val Asn Thr Tyr Ala Gly Asp Leu Gln Lys Lys Leu
65 70 75 80

Val Pro Phe Ala Thr Glu Leu His Glu Arg Leu Ala Lys Asp Ser Glu
85 90 95

Lys Leu Lys Glu Glu Ile Gly Lys Glu Leu Glu Glu Leu Arg Ala Arg

100 105 110

Leu Leu Pro His Ala Asn Glu Val Ser Gln Lys Ile Gly Asp Asn Leu
115 120 125

Arg Glu Leu Gln Gln Arg Leu Glu Pro Tyr Ala Asp Gln Leu Arg Thr
130 135 140

Gln Val Asn Thr Gln Ala Glu Gln Leu Arg Arg Gln Leu Thr Pro Tyr
145 150 155 160

Ala Gln Arg Met Glu Arg Val Leu Arg Glu Asn Ala Asp Ser Leu Gln
165 170 175

Ala Ser Leu Arg Pro His Ala Asp Glu Leu Lys Ala Lys Ile Asp Gln
180 185 190

Asn Val Glu Glu Leu Lys Gly Arg Leu Thr Pro Tyr Ala Asp Glu Phe
195 200 205

Lys Val Lys Ile Asp Gln Thr Val Glu Glu Leu Arg Arg Ser Leu Ala
210 215 220

Pro Tyr Ala Gln Asp Thr Gln Glu Lys Leu Asn His Gln Leu Glu Gly
225 230 235 240

Leu Thr Phe Gln Met Lys Lys Asn Ala Glu Glu Leu Lys Ala Arg Ile
245 250 255

Ser Ala Ser Ala Glu Glu Leu Arg Gln Arg Leu Ala Pro Leu Ala Glu
260 265 270

Asp Val Arg Gly Asn Leu Arg Gly Asn Thr Glu Gly Leu Gln Lys Ser
275 280 285

Leu Ala Glu Leu Gly Gly His Leu Asp Gln Gln Val Glu Glu Phe Arg
290 295 300

Arg Arg Val Glu Pro Tyr Gly Glu Asn Phe Asn Lys Ala Leu Val Gln
305 310 315 320

Gln Met Glu Gln Leu Arg Thr Lys Leu Gly Pro His Ala Gly Asp Val
325 330 335

Glu Gly His Leu Ser Phe Leu Glu Lys Asp Leu Arg Asp Lys Val Asn
340 345 350

Ser Phe Phe Ser Thr Phe Lys Glu Lys Glu Ser Gln Asp Lys Thr Leu

355 360 365

Ser Leu Pro Glu Leu Glu Gln Gln Gln Glu Gln His Gln Glu Gln Gln
370 375 380

Gln Glu Gln Val Gln Met Leu Ala Pro Leu Glu Ser
385 390 395

<210> 4
<211> 15
<212> DNA
<213> Homo sapien

<400> 4
gccaccaytt ctaac 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
ggccctgrtg gctgt 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
agtgtgtrgt ggtga 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapien

<400> 7
ggaggagytg agggc 15

<210> 8
<211> 15
<212> DNA

<213> Homo sapien

<400> 8
ctgagggycc ggctg 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapien

<400> 9
gggacaamct gcgag 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
ccagaacrtg gagga 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapien

<400> 11
agaacgtrga ggagc 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapien

<400> 12
accagcargt ggagg 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapien

<400> 13

ggtcaackcc ttctt 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
ggaacagcca ccayt 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
cgataggtta gaart 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
gaccctggcc ctgrt 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
ccggcgacag ccayc 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
gactacagtg tgtrg 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
tccccgtcac cacya 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapien

<400> 20
ggagctggag gagyt 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
agccgggccc tcarc 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapien

<400> 22
gaggagctga gggyc 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapien

<400> 23
gggcagcagc cggrc 15

<210> 24
<211> 15
<212> DNA

<213> Homo sapien

<400> 24
agatcgggga caamc 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapien

<400> 25
gaagctctcg cagkt 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapien

<400> 26
gatcgaccag aacrt 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapien

<400> 27
ttgagctcct ccayg 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapien

<400> 28
tcgaccagaa cgtrg 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapien

<400> 29

cсttgagctc ctcya 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapien

<400> 30
acctggacca gcarg 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapien

<400> 31
ggaactcctc cacyt 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapien

<400> 32
ggacaaggtc aackc 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapien

<400> 33
gtgctgaaga aggmg 15

<210> 34
<211> 10
<212> DNA
<213> Homo sapien

<400> 34
acagccacca 10

<210> 35
<211> 10
<212> DNA
<213> Homo sapien

<400> 35
taggttagaa 10

<210> 36
<211> 10
<212> DNA
<213> Homo sapien

<400> 36
cctggccctg 10

<210> 37
<211> 10
<212> DNA
<213> Homo sapien

<400> 37
gcgacagcca 10

<210> 38
<211> 10
<212> DNA
<213> Homo sapien

<400> 38
tacagtgtgt 10

<210> 39
<211> 10
<212> DNA
<213> Homo sapien

<400> 39
ccgtcaccac 10

<210> 40
<211> 10
<212> DNA

<213> Homo sapien

<400> 40
gctggaggag 10

<210> 41
<211> 10
<212> DNA
<213> Homo sapien

<400> 41
cgggccctca 10

<210> 42
<211> 10
<212> DNA
<213> Homo sapien

<400> 42
gagctgaggg 10

<210> 43
<211> 10
<212> DNA
<213> Homo sapien

<400> 43
cagcagccgg 10

<210> 44
<211> 10
<212> DNA
<213> Homo sapien

<400> 44
tcggggacaa 10

<210> 45
<211> 10
<212> DNA
<213> Homo sapien

<400> 45

gctctcgcag 10

<210> 46
<211> 10
<212> DNA
<213> Homo sapien

<400> 46
cgaccagaac 10

<210> 47
<211> 10
<212> DNA
<213> Homo sapien

<400> 47
agctcctcca 10

<210> 48
<211> 10
<212> DNA
<213> Homo sapien

<400> 48
accagaacgt 10

<210> 49
<211> 10
<212> DNA
<213> Homo sapien

<400> 49
tgagctcctc 10

<210> 50
<211> 10
<212> DNA
<213> Homo sapien

<400> 50
tggaccagca 10

<210> 51
<211> 10
<212> DNA
<213> Homo sapien

<400> 51
actcctccac 10

<210> 52
<211> 10
<212> DNA
<213> Homo sapien

<400> 52
caaggtcaac 10

<210> 53
<211> 10
<212> DNA
<213> Homo sapien

<400> 53
ctgaagaagg 10

<210> 54
<211> 18
<212> DNA
<213> Homo sapien

<400> 54
tgtaaaacga cggccagt 18

<210> 55
<211> 19
<212> DNA
<213> Homo sapien

<400> 55
aggaaacagc tatgaccat 19

<210> 56
<211> 3613
<212> DNA

<213> Homo sapien

<220>
<221> allele
<222> (571)
<223> PS1: Polymorphic base G or A

<220>
<221> allele
<222> (698)
<223> PS2: Polymorphic base C or T

<220>
<221> allele
<222> (858)
<223> PS3: Polymorphic base G or C

<220>
<221> allele
<222> (908)
<223> PS4: Polymorphic base G or A

<220>
<221> allele
<222> (1042)
<223> PS5: Polymorphic base G or A

<220>
<221> allele
<222> (1449)
<223> PS6: Polymorphic base A or G

<220>
<221> allele
<222> (1559)
<223> PS7: Polymorphic base T or C

<220>
<221> allele
<222> (1560)
<223> PS8: Polymorphic base G or A

<220>
<221> allele
<222> (1700)
<223> PS9: Polymorphic base C or T

<220>

<221> allele
<222> (2355)
<223> PS10: Polymorphic base C or T

<220>
<221> allele
<222> (2361)
<223> PS11: Polymorphic base T or C

<220>
<221> allele
<222> (2464)
<223> PS12: Polymorphic base C or T

<220>
<221> allele
<222> (2471)
<223> PS13: Polymorphic base C or T

<220>
<221> allele
<222> (2520)
<223> PS14: Polymorphic base C or A

<220>
<221> allele
<222> (2579)
<223> PS15: Polymorphic base A or G

<220>
<221> allele
<222> (2589)
<223> PS16: Polymorphic base C or T

<220>
<221> allele
<222> (2620)
<223> PS17: Polymorphic base G or T

<220>
<221> allele
<222> (2719)
<223> PS18: Polymorphic base G or A

<220>
<221> allele
<222> (2721)
<223> PS19: Polymorphic base G or A

<220>
<221> allele
<222> (2919)
<223> PS20: Polymorphic base C or T

<220>
<221> allele
<222> (2985)
<223> PS21: Polymorphic base C or T

<220>
<221> allele
<222> (3027)
<223> PS22: Polymorphic base G or A

<220>
<221> allele
<222> (3036)
<223> PS23: Polymorphic base G or A

<220>
<221> allele
<222> (3147)
<223> PS24: Polymorphic base G or T

<220>
<221> allele
<222> (3196)
<223> PS25: Polymorphic base T or G

<220>
<221> allele
<222> (3238)
<223> PS26: Polymorphic base A or T

<220>
<221> allele
<222> (3279)
<223> PS27: Polymorphic base T or G

<400> 56
tctagagatt taaagtatgt gggaggctgt gcttaggtta tatgtaaata ctatgccatt 60
ttatatcaag gacttgaaca tccatggatt ttggtatctg cagagggtcc tggaatcatt 120
tcccatggag actgagagat gaccgtacta cccacttcgc aagcaatgtc ttctttaatg 180
tactgaacca tcccattgtt cagaggagaa actgaagctc agggctttga ataactagac 240
caaggaggca cagcatggga gtgggacatg aagcactcta caattaaccc tttcaggaca 300
aggccctgtc tcccacaccc atctgcccaa aggctctcca gggccccctc ctcttgggtg 360

taccttgaca agagacctag attttagctc actatgctgt ctgcagtcct ggatggtccc 420
actccagtgt ctggtgctct gagatggagt cagcattagt ggcggatgtg gagactgggg 480
ggacctgtct tcactggggt agacagagga gatgtggact ttgcccccca tgagcccggc 540
acaaacccga gaggccgcca gcagggcctc raggcatcag tcccgggctc atgggctccc 600
tgaggtgttt ctcctactgt tttccgttcc cctcctccct tccatgctga ggttggtggg 660
gtgggggtgg gggtgcccac gcacggaaca gccaccaytt ctaacctatc gcctgagccc 720
tgatctgctg tcagcttcca cgtagtctca gggtcacaaa agtccaagag gcctcttggg 780
aatgtgtcac cttccagcgt ggagtcacac tgaggaagga ggaggggagg gcagccaggg 840
gggtggcgat agggagasag tttaaatgtc tggctggctc tgagcttcag tcagttccca 900
ctgcagcrca ggtgagctct cctgaggacc tctctgtcag ctcccctgat tgtagggagg 960
atccagtgtg gcaagaaact cctccagccc agcaagcagc tcaggatgtt cctgaaggcc 1020
gtggtcctga ccctggccct grtggctgtc gccggtgagt agaagctgtc tttggatggc 1080
actcctgggc tgctgctctg agtagtgcag gatggaggct gagccaaagc aaaaggacac 1140
ttctgagtgc ccatcagccc ccagctggac atgaggtctg cctggctgcc aagtggctca 1200
caggagagct ggcccagtcc cagtggtggg cccattggca ttggtgctat accagtttca 1260
catatccctg tggcttccaa aaagctaagc tcagacaggg aaaatggcag gttgaggcac 1320
ccccaccatc atccagtctg cagctcagag ctggagcaga ggggccacac aggagacggg 1380
gcctcatgaa ttgctctctg ttaccaccca ggagccaggg ctgaggtcag tgctgaccag 1440
gtggccacrg tgatgtggga ctacttcagc cagctgagca acaatgccaa ggaggccgtg 1500
gaacatctcc agaaatctga actcacccag caactcaagt aagagggact acagtgtgyr 1560
gtggtgacgg ggaattctta aaggccatgc aatgtactgg caagggttga gcttagagac 1620
aggagccctg agcttaggat acccactgcc ctgccactaa ctggccgggc ctctgaacct 1680
aggatccaca tatgtaaacy ggaagtttgg accgaataat ccctgccatg tccttttgct 1740
ttgacgttct agagtttgac aaatggccac atcctatcat tcaggctcat ggaagagagg 1800
gagggaggaa aatgtcatgt gagctgattt ctaatacgtt tcagaaagac aggccccagt 1860
ggaatcaagg ggagggaggt gggaatattt gggaggcccc tgggcacagg caaggaaagc 1920
agcaccttgt gccactggaa gaccccagca gaggtcaaga agacaacatt gtgttacaca 1980
atgtgatcct atggcccaga acattccctc tgggaaggac ctcaaagtcc cacccctctgc 2040
agacaaggag gggaaagcaa actgctggag gtgacatggt gggtagattc tgagacaaac 2100
tatgtgggag atcctgagat agaaattcag catcgtaact tagtctgtga cacccatcct 2160
ctccaatctg caccaccata gggagggtga actcggtacc tctgagcact cacctgtctt 2220
agcacgtgtg cataaggcga gtggtataca agcagacaaa gtcttgccgt gtaaatgcca 2280
aatgtaacgt ggcctccttg tgcccttccc cacagtgccc tcttccagga caaacttgga 2340
gaagtgaaca cttaygcagg ygacctgcag aagaagctgg tgccctttgc caccgagctg 2400
catgaacgcc tggccaagga ctcggagaaa ctgaaggagg agattgggaa ggagctggag 2460
gagytgaggg yccggctgct gccccatgcc aatgaggtga gccagaagat cggggacaam 2520
ctgcgagagc ttcagcagcg cctggagccc tacgcggacc agctgcgcac ccaggtcarc 2580
acgcaggcyg agcagctgcg gcgccagctg acccccctack cacagcgcat ggagagagtg 2640
ctgcgggaga acgccgacag cctgcaggcc tcgctgaggc cccacgccga cgagctcaag 2700
gccaagatcg accagaacrt rgaggagctc aagggacgcc ttacgcccta cgctgacgaa 2760
ttcaaagtca agattgacca gaccgtggag gagctgcgcc gcagcctggc tccctatgct 2820
caggacacgc aggagaagct caaccaccag cttgagggcc tgaccttcca gatgaagaag 2880
aacgccgagg agctcaaggc caggatctcg gccagtgcyg aggagctgcg gcagaggctg 2940
gcgcccttgg ccgaggacgt gcgtggcaac ctgaggggca acacygaggg gctgcagaag 3000
tcactggcag agctgggtgg gcacctrgac cagcargtgg aggagttccg acgccgggtg 3060
gagccctacg gggaaaactt caacaaagcc ctggtgcagc agatggaaca gctcaggacg 3120
aaactgggcc cccatgcggg ggacgtkgaa ggccacttga gcttcctgga gaaggacctg 3180
agggacaagg tcaackcctt cttcagcacc ttcaaggaga aagagagcca ggacaagwct 3240

```
ctctccctcc ctgagctgga gcaacagcag gaacagcakc aggagcagca gcaggagcag 3300
gtgcagatgc tggccccttt ggagagctga gctgcccctg gtgcactggc cccaccctcg 3360
tggacacctg ccctgccctg ccacctgtct gtctgtctgt cccaaagaag ttctggtatg 3420
aacttgagga cacatgtcca gtgggaggtg agaccacctc tcaatattca ataaagctgc 3480
tgagaatcta gcctcaactg gttgccggat gaatcctcct tgcagctggg gaggtgggga 3540
ggtaaccatg actgggcaga gcttagcagc ggcctggcag gagacaccca ggattgggga 3600
gatgagactg cag                                                   3613
```

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,013,431
DATED : January 11, 2000
INVENTOR(S) : Soderlund et al.

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Column 17,
Line 16, after "[35", second occurrence, insert -- S --.
Line 36, after "[35", insert -- S --.
Line 40, after "[3", insert -- H --, after "]", delete -- H --.
Line 49, after "[35", insert -- S --, replace "p" with -- P --.
Line 54, after "[", insert -- 32 --, after "P", delete -- 32 --.
Line 56, after "[35", insert -- S --.
Line 57, replace "p" with -- P --.

Signed and Sealed this

Eleventh Day of September, 2001

Attest:

Nicholas P. Godici

NICHOLAS P. GODICI

Attesting Officer *Acting Director of the United States Patent and Trademark Office*